

INFORMATION EXTRACTION FROM DOCUMENTS WITH REGULAR EXPRESSION MATCHING

Abstract of the Disclosure

Techniques are provided for enumerating regularly identifiable or stereotypical phrases that people commonly use to convey particular information, and where exactly in these phrases the particular information is to be found. In one embodiment, such phrases are referred to as "regular expressions." Using such enumerated phrases, the invention is able to automatically identify them in an input data stream and then identify and extract the particular information associated with the phrase that is being sought, e.g., important or relevant information.

1500-211.APP